



**QUEEN'S
UNIVERSITY
BELFAST**

The development of time estimation: Comparing childhood and old age.

McCormack, T., Brown, G. D. A., Maylor, E. A., Darby, R. J., & Green, D. (1999). The development of time estimation: Comparing childhood and old age. *Developmental Psychology*, 35, 1143-1155.

Published in:

Developmental Psychology

Document Version:

Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Developmental Changes in Time Estimation: Comparing Childhood and Old Age

Teresa McCormack, Gordon D. A. Brown, Elizabeth A. Maylor, Richard J. Darby, and Dina Green
University of Warwick

Participants from ages 5 to 99 years completed 2 time estimation tasks: a temporal generalization task and a temporal bisection task. Developmental differences in overall levels of performance were found at both ends of the life span and were more marked on the generalization task than the bisection task. Older adults and children performed at lower levels than young adults, but there were also qualitative differences in the patterns of errors made by the older adults and the children. To capture these findings, the authors propose a new developmental model of temporal generalization and bisection. The model assumes developmental changes across the life span in the noisiness of initial perceptual encoding and across childhood in the extent to which long-term memory of time intervals is distorted.

Recent theorizing in comparative psychology has suggested that the kinds of basic mechanism that underpin certain animal timing behaviors, such as lever pressing at fixed temporal intervals, may also be involved in human duration estimation (e.g., Allan & Gibbon, 1991; Wearden 1994; Wearden & Lejeune, 1993). Wearden (1994) argued that human performance on some simple timing tasks exploits *biopsychological time*, defined as time experience “based directly on (hypothetical or real) biologically-based timing mechanisms such as internal clocks or oscillators” (Wearden, 1994, p. 217). Much progress has been made in uncovering the nature of these processes shared by humans and animals, as well as their neural basis (Brown & Vousden, 1998; Church, 1984; Meck, Church, & Olton, 1984; Olton, 1989; Wearden, 1991a). Some of the changes over the life span in time estimation (for reviews, see Block, Zakay, & Hancock, 1998, 1999) could be due to the development of these basic timing mechanisms (Wearden, Wearden, & Rabbitt, 1997).

Our aim in this article is to examine whether there are developmental changes across the life span in biopsychological timing and whether such changes can be explained within the *scalar timing* framework. A number of well-specified scalar timing models have emerged that can account for a broad range of animal and human data (see Gibbon, Church, & Meck, 1984, and Wearden, 1991a, 1995, for accounts of scalar timing theory). According to

scalar timing theory, timing behavior is based on the output of an internal clock that provides long-term memory representations that can be retrieved and compared with the representation of the current temporal interval held in short-term memory. Given that there are a number of different parameters that can vary in such models (e.g., in the clock or memory processes), it is possible to generate developmental versions of the models (Wearden, Wearden, et al., 1997). However, as far as we are aware, developmental changes in timing across childhood have not been considered within this framework.

Wearden and his colleagues have developed new methodologies for examining human stimulus timing that are based directly on paradigms in the animal timing literature. Two tasks in particular have received attention: the temporal generalization task and the temporal bisection task (Wearden, 1991b, 1992; Wearden & Ferrara, 1995, 1996). In the temporal generalization task, participants are exposed to a standard stimulus of a fixed duration. They are then required to judge whether subsequently presented stimuli are the same duration as the standard stimulus. In the temporal bisection task, participants are exposed to two standard durations, one long and one short. They then must judge whether presented durations are more similar to the long or the short standard. Although the versions of these tasks used with humans differ in a number of ways from the animal tasks (e.g., in terms of the kind of feedback given to participants and the typical lengths of the stimuli), they are structurally similar. Studies using these tasks have provided a rich source of data that have enabled meaningful human–animal comparisons in timing and enhanced our understanding of the nature of timing judgments in humans (Wearden, in press; Wearden & Lejeune, 1993; Wearden & McShane, 1988).

Despite the success of these paradigms, developmental changes on such tasks have only been examined in a single study, that of Wearden, Wearden, et al. (1997). They found that older participants (between ages 70 and 79 years) showed a decline in performance on a temporal generalization task, compared with both a younger elderly group (between ages 60 and 69 years) and an undergraduate group. However, Wearden, Wearden, et al. found no significant age differences on a version of the bisection task, which they suggested may be a less sensitive task.

Teresa McCormack, Gordon D. A. Brown, Elizabeth A. Maylor, Richard J. Darby, and Dina Green, Department of Psychology, University of Warwick, Coventry, England.

This research was supported by grants from the Research Councils of the United Kingdom (Medical Research Council Grants G9608199 and G9606610N and Economic and Social Research Council Grant R 000 23 2576). We are very grateful to Jemma Rosen-Webb for help in testing participants and to Fiona Anderson for recruiting participants and scoring data in Experiment 2. John Wearden provided helpful advice on the design of the study.

Correspondence concerning this article should be addressed to Teresa McCormack, Department of Psychology, University of Warwick, Coventry CV4 7AL, England. Electronic mail may be sent to T.McCormack@warwick.ac.uk.

Although the bisection and generalization tasks only measure certain aspects of timing behavior (timing of short, unfilled intervals), they have a number of methodological advantages that are particularly relevant in a developmental context. First, the stimulus durations that are used are short enough (typically less than 1 or 2 s) to prevent chronometric counting. This is especially important from a developmental perspective, because it ensures that developmental differences are not due to changing competence in counting. It should be noted that in adults similar patterns of findings on these tasks are found with longer durations, providing counting is suppressed, which suggests that the tasks measure some general characteristics of human timing (Wearden, Denovan, Fakhri, & Haworth, 1997). Second, both of these tasks involve making simple comparative judgments about durations (same-different judgments in the case of temporal generalization, and more-similar-to judgments in the bisection task). This ensures that developmental differences on such tasks are not due to changes in the ability to use a scale, as required in some temporal estimation tasks, or changes in the ability to inhibit or initiate a response, as may be required in production tasks. Third, these tasks have been shown to produce consistent and orderly data in young adult populations across a wide variety of experimental conditions and therefore provide a good background for interpretation of developmental findings. In particular, the data generated from these tasks can be interpreted within the general framework of scalar timing theory (Gibbon, 1977).

In the present study, modified versions of the generalization and bisection tasks were used to examine changes in timing across a broad range of the life span, from childhood to old age. We were particularly interested in examining whether developmental differences in timing behavior in childhood are qualitatively similar to developmental differences in an aging population.

Experiment 1: Temporal Generalization

The human temporal generalization procedure is based on a paradigm originally used by Church and Gibbon (1982) to study temporal discrimination in rats. In Church and Gibbon's study, rats were presented with a series of nine visual signals (typically ranging from approximately 1 to 8 s) after having been reinforced for responding to the middle stimulus in the series (the standard) but not for responding to the signals that were shorter or longer than the standard. Plotting the probability of responding to a given signal in the series against signal duration yields a *generalization gradient* that peaks at the standard and declines with distance from it. Wearden (1992) showed that orderly data could be obtained from humans using a very similar task that involved identifying a standard stimulus of a given length from a series of seven stimuli. To prevent the use of counting strategies, Wearden used stimuli that were much shorter than those used with rats (typically between 100 and 700 ms), although Wearden, Denovan, et al. (1997) subsequently obtained similar functions using much longer durations.

Although scalar timing models can fit both the human and animal data extremely well, a crucial difference lies in the symmetry of the generalization gradients. Those obtained with rats are usually approximately symmetrical bell-shaped distributions such that the probability of responding is similar for durations at equal distances above and below the standard. However, Wearden has

consistently found that human generalization gradients are asymmetrical in real time, with positive responses more likely to stimuli longer than the standard than to those shorter than the standard but of equal distance away. He has accounted for the difference between the shapes of animal and human generalization gradients by proposing that the decision rule used by humans differs from that used by rats (e.g., Wearden, 1992).

These considerations suggest that across different populations, not only may the steepness of the temporal generalization gradient change (indicating a change in overall level of performance), but symmetry of the gradient may also vary. In the present study, we adapted Wearden's human temporal generalization procedure to compare patterns of development in childhood and old age directly. The most notable difference between our version of the task and that used by Wearden, Wearden, et al. (1997) is that we placed our stimulus identification task in a context that could be more easily understood by young children (and perhaps elderly adults). Whereas Wearden, Wearden, et al.'s participants were explicitly instructed to try to identify a standard length tone, in our task participants were told that the standard was the sound that belonged to a bird pictured on a computer screen. They then had to judge whether other sounds of different lengths were the bird's sound or not. Children were able to give their response by pointing at one of two visual displays rather than by pressing a key or giving a verbal response.

Method

Participants. Two samples of participants took part in the study and were tested at different times. The first sample had four groups of participants: 26 five-year-olds ($M = 5.7$ years, range = 5.3 to 6.8 years; 14 girls and 12 boys), 32 eight-year-olds ($M = 8.6$ years, range = 8.2 to 9.1 years; 13 girls and 19 boys), 34 ten-year-olds ($M = 10.8$ years, range = 10.3 to 11.2 years; 18 girls and 16 boys), and 26 young adults ($M = 19.1$ years, range = 18.3 to 21.7 years; 22 women and 4 men). Testing of all of the participants in this sample was carried out individually. Children were tested in their schools and were given a small gift for participation. The young adults were undergraduate students at the University of Warwick who received course credit for participation: In what follows, this group will be referred to as the undergraduate group.

The second sample was tested approximately 2 months later and did not include any participants from the first sample. The majority of these participants were tested in groups rather than individually. The sample comprised participants of three age ranges: young ($n = 36$; 24 women and 12 men), young-old ($n = 55$; 32 women and 23 men), and old-old ($n = 33$; 17 women and 16 men). The young participants were between ages 16 and 25 years, and the majority were pupils at a local higher education college. All of the young participants were tested in groups and received no payment for their participation in the study. The young-old participants were between ages 63 and 75 years, and the old-old participants were between ages 75 and 99 years. Some of the participants in the young-old and old-old groups ($n = 49$ and $n = 19$, respectively) were recruited through local newspaper articles asking for volunteers to take part in a study of memory and aging; they were required to make their own travel arrangements to attend a group testing session at the University of Warwick and were paid U.K. £5 as a contribution toward their travel expenses. The remaining young-old and old-old participants ($n = 6$ and $n = 14$, respectively) were tested individually in day centers or residential homes for the elderly; they received no payment. Participants in the latter group scored at least 25 out of 30 on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975); thus they were unlikely to have been suffering from any form of dementia.

Table 1
Background Scores for the Young, Young-Old,
and Old-Old Age Groups

Measure	Young (<i>n</i> = 36)		Young-old (<i>n</i> = 55)		Old-old (<i>n</i> = 33)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age (in years)	19.6	2.2	69.6	3.2	81.5	5.8
Fluid ability ^a	—	—	68.2 ^b	15.9 ^b	57.5 ^c	19.3 ^c
Vocabulary ^d	17.4	4.2	22.9 ^b	3.7 ^b	22.9 ^c	4.9 ^c
Speed ^e	66.6	11.2	42.7 ^b	10.0 ^b	37.7 ^c	11.6 ^c
Current health ^f	4.0	0.7	3.9 ^b	0.6 ^b	3.7 ^c	0.9 ^c
Eyesight ^f (corrected)	4.3	0.8	4.0 ^b	0.6 ^b	4.0 ^c	0.7 ^c
Hearing ^f (corrected)	4.2	0.8	3.8 ^b	0.7 ^b	3.6 ^c	0.8 ^c

Note. Dashes indicate that data were not obtained.

^a AH4 test, a timed problem-solving test of fluid ability or intelligence (Heim, 1968); maximum score = 130. ^b *n* = 49. ^c *n* = 19. ^d Part 1 of the Mill Hill Vocabulary Test (Raven, Raven, & Court, 1988); maximum score = 33. ^e Digit Symbol Substitution test from the Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981). ^f Self-rated on a 5-point scale (1 = very poor; 2 = poor; 3 = fair; 4 = good; 5 = very good).

The mean ages for each of the three age groups in the second sample are presented in Table 1. Further background information was available for the participants who were tested in groups at the University of Warwick (see Table 1). The AH4 is a timed problem-solving test of fluid ability or intelligence (Heim, 1968), equally divided between verbal and arithmetic problems and spatial problems; the two halves of the test have been combined in Table 1. In the first part of the Mill Hill Vocabulary Test (Raven, Raven, & Court, 1988), participants are required to select the best synonym for a target word from a set of six alternatives. Speed was measured by the Digit Symbol Substitution test from the Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981). Analyses of variance (ANOVAs) revealed significant age differences for fluid ability, $F(1, 66) = 5.44$, $MSE = 285.76$, $p < .05$; vocabulary, $F(2, 101) = 20.70$, $MSE = 17.08$, $p < .0001$; and speed, $F(2, 101) = 67.33$, $MSE = 114.30$, $p < .0001$. For fluid ability, the young-old group outperformed the old-old group. The AH4 test was not administered to the young group; however, the mean score for an equivalent group of 68 undergraduate students with the same mean age of 20 years was 104, that is, considerably higher than the present young-old and old-old groups. For vocabulary, post hoc tests revealed significant improvement from the young group to the young-old group but no difference between the young-old and old-old groups. Speed declined significantly from the young group to the young-old group on a post hoc test, but the decline from the young-old group to the old-old group did not reach significance. Thus, the present background data are consistent with the pattern typical in the aging literature of age-related decline in fluid intelligence and speed but growth in crystallized intelligence as indicated in this case by vocabulary (cf. Horn & Cattell, 1967; Salthouse, 1991).

Self-rated measures of participants' current state of health, eyesight (with glasses, if worn), and hearing (with aids, if worn) are also included in Table 1. These were all generally high, with averages equivalent to ratings of *good*. The age groups did not differ in terms of health, $F(2, 101) = 1.59$, $MSE = 0.46$, $p > .1$, or eyesight, $F(2, 101) = 1.94$, $MSE = 0.47$, $p > .1$. However, there was a significant effect of age group on hearing, $F(2, 101) = 5.05$, $MSE = 0.54$, $p < .01$; post hoc comparisons revealed significantly higher ratings for the young group than for the young-old group, with no difference between the young-old and old-old groups.¹ All participants from both samples took part in Experiment 1 and Experiment 2, with the order in which the tasks were completed counterbalanced.

Apparatus and stimuli. The experiment was run on an Apple Macintosh computer, and stimulus presentations were controlled by the Superlab

software package. The auditory stimuli were 500-Hz tones produced by the computer's speaker, and the visual displays were presented on a black-and-white screen. Children gave their responses to stimuli by pointing to parts of the visual displays, as did those elderly adults who were tested individually in day centers or residential homes. The other adults from the first and second samples gave written responses by circling yes or no on a response sheet.

Procedure. The participants tested in groups (maximum of 10 per group) were seated at individual desks in a small laboratory. The average viewing distance from the computer screen was approximately 2 m, with no participant being more than 3.8 m away from the screen. Those with either poor eyesight or poor hearing were encouraged to sit at the front of the room. Young participants were tested in separate sessions from young-old and old-old participants, who were tested together.

All participants were randomly assigned to one of two versions of the task, with the order in which the stimuli were presented in each trial reversed between versions. Following Wearden, Wearden, et al. (1997), stimuli were presented in eight trials, with each trial consisting of a series of eight stimulus presentations, making a total of 64 experimental stimuli. The standard stimulus duration was 500 ms, and the nonstandard durations were 125, 250, 375, 625, 750, and 875 ms. Each series of eight stimulus presentations contained one example of each of the nonstandard stimuli and two examples of the standard duration.

Participants were told that they were taking part in a task in which they would hear some sounds and have to make judgments about their length. The first phase of the introduction to the experiment was an initial exposure to the standard stimulus. In this phase, a picture of an owl appeared in the center of the screen. Participants were told that the owl always made a sound of the same length and were instructed to listen carefully to how long the sound was. They then heard the owl's 500-ms sound five times, while the owl remained stationary in the center of the screen. They were told that they were going to hear some more sounds and that their task was to judge whether those sounds were the owl's sound.

In the second phase of the introduction, the experimenter demonstrated the task. Two pictures appeared on the screen side by side, one showing the owl and the other showing the owl crossed out, with a question mark situated between the two pictures. A nonstandard duration of 750 ms was played, and the experimenter explained that this was not the owl's sound because it was too long. To the children and to the elderly participants tested individually, the experimenter explained that the correct response was to point to the crossed-out owl and demonstrated this pointing response. To the other adults, the experimenter explained that the correct response was to circle *no* on their response sheet. This was followed by an example of the 500-ms sound, for which the correct response was to point at the owl or to circle *yes* on the response sheet, and then a 250-ms nonstandard duration, with the experimenter demonstrating and explaining the appropriate responses in each case. A practice trial followed this demonstration, in which participants heard one example each of the four nonstandard durations that they had not yet encountered (125, 375, 625, and 875 ms), plus another example of the standard duration. Once a response had been made, participants were informed whether it was correct.

Following the practice trial, participants were told that they would hear more sounds and would have to judge again whether they were or were not the owl's sound. They were also reminded of the nature of the feedback that would be given throughout the task. The last phase of the introduction was a second exposure to the standard duration: The owl's sound was played five more times, with the owl again displayed in the center of the

¹ The age differences in background scores shown in Table 1 are probably underestimates of the age differences across the two older age groups because of the absence of data from the participants in day centers and residential homes who were generally very elderly.

screen. The owl was then replaced by the pictures of the owl and the crossed-out owl, and the first experimental trial began immediately.

During individual testing, the experimenter always sat behind the participant. In order for the task to be paced to suit each participant, the production of each test stimulus was controlled by the experimenter using a key press (unseen by participants). Thus, the delay between feedback and the next auditory event was not of a fixed length of time. The experimenter also controlled the delay between each series of trials in response to the participant's readiness to begin each trial.

For group testing, the experimenter stood at the front of the laboratory and controlled stimulus presentation by pressing a key on the computer keyboard which was hidden from participants. The number of each stimulus was announced and then presented after an interval of approximately 2 s. When all participants had made their responses and were looking up from their response sheets, the experimenter provided the correct response (either "No, that was not the owl" or "Yes, that was the owl").

Results

All of the participants completed the task, and for each participant the proportion of yes responses to each stimulus type was calculated. There were 18 missing data points from those participants who had given written responses, 9 of which came from 1 young-old participant who was therefore deleted from the analysis. The remaining 9 missing data points (from 7 participants) were replaced by each participant's mean for the relevant condition.

Inspection of the data suggested that a number of the children and older adults failed to understand or comply with task instructions. Following Wearden, Wearden, et al. (1997), these participants were excluded from the analysis. Failure to follow task instructions is shown by a flat distribution of positive responses across all stimulus types, rather than a distribution that peaks at or near the standard stimulus. Therefore, participants were excluded if the largest difference in the proportion of positive responses to any two stimuli was less than or equal to 0.5. This was felt to be a conservative criterion that would reduce the likelihood of finding significant age differences. Ten of the 5-year-olds and 2 of the 8-year-olds were excluded from the analysis by this criterion, as were 2 young-old and 6 old-old participants. It should be emphasized that analysis of the full data set yields patterns of develop-

mental differences, including the qualitative differences in performance found between age groups, which are very similar to those reported below. Analyses were carried out separately on the data from the first sample (the children and undergraduates) and the second sample (the young, young-old, and old-old adults), although we report both sets of analyses simultaneously.

Because the task involved forced-choice responding, an initial analysis was carried out to establish whether there were developmental changes in an overall bias to give positive or negative responses. ANOVAs were performed on the overall proportion of yes responses calculated by averaging the proportion of yes responses to standard stimulus durations with the proportion of yes responses to nonstandard stimulus durations. For the children and undergraduates, the effect of age was not significant, $F(3, 102) = 1.46$, $MSE = 0.006$, $p = .23$; however, the effect approached significance for the young and older adults in the second sample, $F(2, 112) = 2.85$, $MSE = 0.008$, $p = .06$. Thus, we decided to normalize the proportions of positive responses for each participant (i.e., for each stimulus type, the number of yes responses was divided by the total number of yes responses across all stimuli). Analysis of the unnormalized proportions yields findings very similar to those that will now be reported.

Figures 1a and 2a illustrate the distribution of yes responses across stimulus durations for each age group. Although generalization gradients were flatter for the children, the young-old, and the old-old groups, all groups produced an orderly pattern of responses, with generalization gradients peaking at the standard duration and the proportion of positive responses declining with distance from the standard. For each sample, an ANOVA on the normalized data with a between-subjects variable of age and a within-subjects variable of stimulus duration was carried out. Each ANOVA yielded a significant effect of stimulus duration: $F(6, 612) = 236.12$, $MSE = 0.009$, $p < .001$, for the children and undergraduates; $F(6, 672) = 160.85$, $MSE = 0.012$, $p < .01$, for the young, young-old, and old-old adults.

Comparison of the gradients for different age groups suggested that age differences did not occur at all stimulus durations, and the ANOVAs confirmed this, because the interactions between age

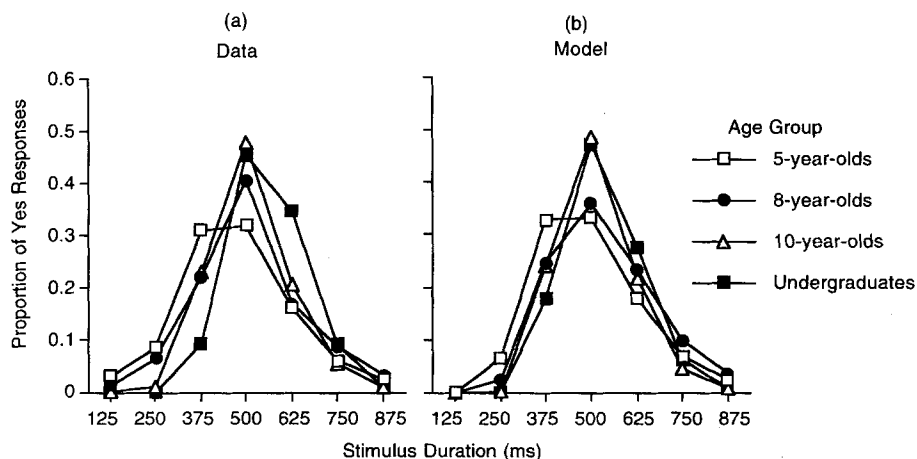


Figure 1. Proportion of yes responses (normalized data) as a function of stimulus duration and age group for the children and undergraduates: (a) the data and (b) the model.

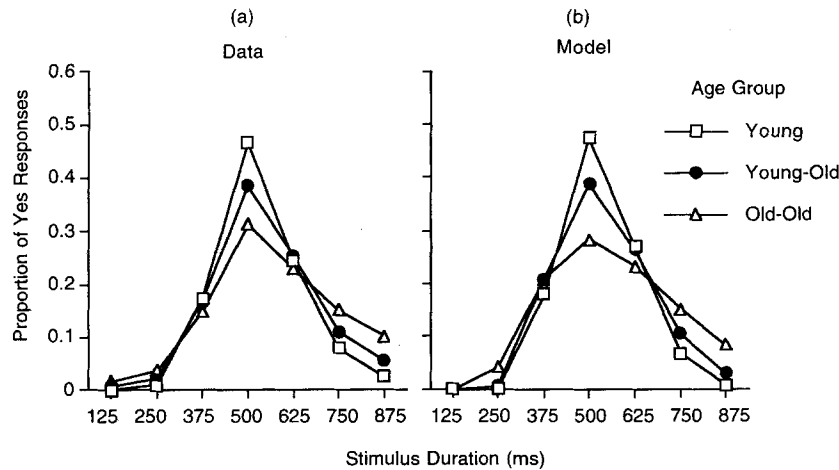


Figure 2. Proportion of yes responses (normalized data) as a function of stimulus duration and age group for the young and older adults: (a) the data and (b) the model.

and stimulus duration were significant: $F(18, 612) = 9.48$, $MSE = 0.009$, $p < .001$, for the children and undergraduates; $F(12, 672) = 3.91$, $MSE = 0.012$, $p < .01$, for the three adult groups from the second sample.² Differences between the children and undergraduates appeared to be more marked for the longer durations, whereas differences between young and older adults were more marked for the shorter durations. Analysis of simple effects confirmed these impressions. Analysis of simple effects on the performance of the groups of young and young-old and old-old adults (Figure 2a) revealed that there were significant age differences at stimulus durations of 500, 750, and 875 ms only: $F(2, 112) > 4.7$, $p < .02$, for all comparisons. Post hoc comparisons were conducted to examine the age differences between each of the adult groups on these three stimulus durations; the only differences to reach significance at the $p < .05$ level were between the young and old-old groups.

By contrast, analysis of simple effects showed that the age differences between the groups of children and undergraduates were significant at durations of 125, 250, 375, 500, and 625 ms, $F(3, 102) > 4.4$, $p < .01$ for all comparisons, but not on the two longest durations of 750 and 875 ms. The absence of age differences on these durations indicates that even 5-year-olds were able to make some temporal discriminations as well as adults, and therefore that they were able to understand and comply with task instructions. Further post hoc tests compared the age differences between each of the groups on the 125-, 250-, 375-, 500-, and 625-ms stimuli (a significance level of $p < .05$ was taken for all post hoc comparisons). The proportion of positive responses to the standard (500-ms) stimulus was only significantly different between the 5-year-olds and the 10-year-olds and undergraduates. None of the differences between the 5-year-olds and the 8-year-olds was significant for any stimuli, but the 5-year-olds produced a significantly higher proportion of positive responses to the 125- and 250-ms stimuli than the undergraduates or the 10-year-olds. Undergraduates produced a significantly higher proportion of positive responses to the 625-ms stimulus and a significantly lower proportion of positive responses to the 375-ms stimulus than did any of the three age groups of children, suggesting a qualitatively

different pattern of responding in undergraduates to that in children. The three groups of children did not differ significantly from each other on these two stimuli.

Studies of temporal generalization in adults have consistently found right asymmetry in the generalization gradient (i.e., it is skewed to the right), with more positive responses to stimulus durations longer than the standard than to those shorter than the standard but of equal distance from it. Inspection of the generalization gradients suggested that right asymmetry was present in all the adult groups. However, the 5-year-olds and 8-year-olds showed left asymmetry, whereas the gradient of the 10-year-olds appeared approximately symmetrical. Asymmetry was tested by comparing the proportion of yes responses to each of the three pairs of durations of equal distance from the standard, with the closest pair being defined as the pair one step (125-ms) from the standard (325 and 625 ms), the pair two steps from the standard (250 and 750 ms) being labeled the middle pair, and the furthest pair being that three steps from the standard (125 and 875 ms). Paired t tests found significant differences in the middle and closest pairs for all the adult groups and in the middle pair for the 10-year-old group, with more positive responses being given to the longer stimulus in each pair. The members of the closest pair were also significantly different in the 5-year-old age group, but in the 5-year-old group the proportion of positive responses to the 375-ms stimulus was significantly greater than that to the 625-ms stimulus, whereas this difference was reversed in the adult groups.

Thus, whereas all of the adults showed significant right asymmetry, the youngest age group showed significant left asymmetry. The 8- and 10-year-olds showed less asymmetry than did the adult groups or the 5-year-olds, although inspection of the gradients indicates that all groups of children show a general trend for left asymmetry. Therefore, a developmental trend from significant left asymmetry to significant right asymmetry was in evidence.

² In the ANOVAs with repeated measures, in which there was evidence of departure from the sphericity assumption, the reported probability levels have been adjusted accordingly (Greenhouse-Geisser corrections).

Last, we examined the relationships between the young, young-old, and old-old groups' performance on the temporal generalization task and processing speed and IQ. For these analyses, a single measure of the proportion correct was used (the number of yes responses to the 500-ms stimulus divided by the total number of yes responses). As expected from the results already presented, there was a significant correlation between age and performance, $r(113) = -.335$, $p < .001$. A measure of processing speed (the Digit-Symbol Substitution test from the Wechsler Adult Intelligence Scale—Revised; Wechsler, 1981) was available for the majority of participants in the young, young-old, and old-old groups. The correlation between temporal generalization and processing speed was also significant, $r(98) = .224$, $p < .05$. Crucially, the partial correlation between age and performance, with the influence of processing speed partialled out, was not significant, $r(98) = -.102$, $p > .1$. In other words, the age-related decline in temporal generalization is consistent with the reduced-processing-speed theory of cognitive aging (Salthouse, 1996).

Wearden, Wearden, et al. (1997) obtained a significant correlation of .258 between temporal generalization and fluid ability scores from Scale 2 of the Culture Fair Intelligence Test (Cattell & Cattell, 1960). In the present study, for the reduced sample of young-old and old-old participants for whom fluid ability scores were available from the AH4 test (Heim, 1968), this correlation was also positive but did not reach significance, $r(62) = .158$, $p > .1$.

Discussion

As in Wearden, Wearden, et al.'s (1997) study, changes in level of performance with age were found on the temporal generalization task. In this respect, our results are compatible with those of Wearden, Wearden, et al. despite the differences in the task procedures. Also consistent with Wearden, Wearden, et al.'s findings, we found that the overall decline in performance with old age was not accompanied by qualitative changes in the shape of the generalization gradient. By contrast, in addition to overall differences in level of performance between children and undergraduates, there were also qualitative differences in the shape of the generalization gradient. Although all of the adult groups in our study showed marked right asymmetry, the 5-year-old group showed significant asymmetry in the opposite direction, a tendency that declined with age but was also evident in the 8- and 10-year-olds.

Thus, the children's performance differed qualitatively from adult human performance and was also different from the performance of animals on a similar task, with animals tending to produce symmetrical gradients. Although the difference in symmetry between human and animal gradients has been accounted for in terms of differences in decision rules (Wearden, 1992), there is no plausible alteration of such rules that could capture the developmental shift from left asymmetry to right asymmetry. The only previously reported finding of left asymmetry in human temporal generalization that we are aware of was in a study by Penton-Voak, Edwards, Percival, and Wearden (1996, Experiment 1), where it was accounted for in terms of a systematic mismatch between the time representations encoded in long-term memory of the standard and the representations of subsequent intervals. In that task, an arousal manipulation was used to increase the speed of the internal clock after initial exposure to the standard.

Penton-Voak et al. argued that this resulted in the subsequently presented intervals being encoded as longer than they actually were; therefore, intervals that were shorter than the standard tended to be confused with it, giving left asymmetry in the generalization gradient.

In the present task, no such arousal manipulation was used, and there is no reason to suppose that, in children, the encoding of the presented intervals differed from the initial encoding of the standard interval. However, a similar pattern of findings would be predicted if the representations in long-term memory of the standard became distorted (rather than the representations of subsequent stimuli, as in the Penton-Voak et al., 1996, study) such that it was systematically remembered as being shorter than it actually was. If the extent of this shortening distortion changes with age (while the decision rule remains unchanged), then a developmental shift from significant left asymmetry to significant right asymmetry is predicted. The suggestion that the memory representation of the standard may distort in some way in particular populations has been made before in the context of aging in rats by Lejeune, Ferrara, Soffie, Bronchart, and Wearden (1998), who pointed out that experimental manipulations, such as the administration of drugs, can also have this effect (Meck, 1996).

Experiment 2: Temporal Bisection

The bisection task differs from the generalization task in that there are two standards, one long and one short. The task is to decide whether presented durations are more similar to the long or the short standard. In the original version of the task (Church & Deluty, 1977), rats were trained to respond on one lever in response to a short signal and on another lever in response to a different, long signal (typically four times longer) until they showed high discrimination of the two signals. In the test phase, rats were presented with a series of stimuli of durations ranging between and including the two standard durations, and their choice of lever was recorded. The proportion of responses appropriate to the long signal was plotted against stimulus duration, giving a characteristic S-shaped function. One important measure of timing performance given by such plots is the bisection point, which is the duration at which 50% of responses are those appropriate to the long signal. Typically, in studies with rats and pigeons, this point has been found to lie at the geometric mean of the two standard durations (Church & Deluty, 1977; Maricq, Roberts, & Church, 1981; Meck, 1983) rather than at the arithmetic mean (i.e., at the point that is the square root of the product of the two standards rather than the point midway between them on a linear scale). The location of this point has implications for scalar timing theory, such as whether subjective time is considered to be linearly or logarithmically spaced or what kinds of decision processes are thought to act on time representations (Gibbon, 1981; Killeen & Fetterman, 1988; Raslear, 1983; Wearden & Ferrara, 1995).

A number of studies have shown that it is possible to adapt this procedure for use with humans (Allan & Gibbon, 1991; Wearden, 1991b; Wearden & Ferrara, 1995, 1996). The human version of the task is very similar and involves judging whether a series of stimulus durations are more similar to a long or a short standard. As with the human temporal generalization procedure, the durations used are often considerably shorter than those used with animals, although durations up to 10 s have been used along with

counting suppression. Although this procedure yields orderly S-shaped bisection plots in humans, the bisection point has not consistently been found to be located at or near the geometric mean, unlike in animal studies. Although Allan and Gibbon (1991) found bisection close to the geometric mean with linearly spaced stimulus durations, Wearden's studies have tended to show bisection closer to the arithmetic mean across a variety of stimulus ranges (see Wearden, Rogers, & Thomas, 1997). Wearden has suggested that a variety of findings of bisection at the arithmetic mean can be accounted for by assuming that the decision processes used by humans on this task are very similar to those used in the temporal generalization task (Wearden & Ferrara, 1995).

As with temporal generalization, these considerations of the qualitative differences between animal and human performance suggest that there may be population differences in the location of the bisection point as well as in the overall level of performance. Our version of the bisection task, like the generalization task, was very similar to that used in Wearden, Wearden, et al.'s (1997) study, although again the procedure was altered to enable young children and the elderly adults to understand the task. The long and the short standards were initially presented as the sounds of a big bird and a small bird, which were displayed visually on a computer screen. Participants had to choose whether subsequent sounds were more similar in length to the big bird's sound or to the small bird's sound. Wearden, Wearden, et al. found that the bisection task was less sensitive to age and suggested that this may have been due to the frequent reminders that participants received of the standard stimuli between trials. In our version of the task, we removed these reminders to examine whether this would increase the developmental sensitivity of the task.

Method

Participants. These were the same two samples that had participated in Experiment 1.

Apparatus and stimuli. The same apparatus that was used in the temporal generalization task was used in this experiment. As before, the stimuli were 500-Hz tones played over the computer's speaker.

Procedure. The short standard stimulus duration was 200 ms and the long standard was 800 ms, with the nonstandard stimulus durations being 300, 400, 500, 600, and 700 ms. Participants received five trials consisting of a series of seven stimulus presentations, one presentation of each of the seven stimuli. The order in which the stimuli was presented was fixed for half of the participants, with the other half assigned to a condition in which the presentation order in each trial was reversed.

The testing conditions were identical to those used in the temporal generalization experiment, with children and a minority of older adults tested individually and all other adults tested in groups. In the initial exposure phase, participants were shown a display in which two birds, one small and one big, appeared side by side on the screen. They were informed that the birds made sounds of different lengths, with the small bird making a short sound and the large bird making a long sound. They were then given five alternating presentations of the short and long sounds, with the appropriate bird appearing on the screen during the presentation of each sound. Participants were then told that they had to judge whether some other sounds were more like the small bird's sound or more like the big bird's sound in terms of their length. A practice series of the seven stimulus durations then followed, during which both of the birds were displayed side by side on the screen.

Children and the older adults who were tested individually gave their response by pointing to the appropriate bird, and the other adults gave their response by circling either *small* or *big* on their response sheet. The five

test series followed immediately (i.e., 35 experimental stimuli in total). No feedback was given in this task, and further reminders of the short and long standards were not given.

Results

All of the participants completed the task, and the proportion of long (i.e., *big*) responses to each stimulus duration was calculated for each participant. There was only one missing data point; this was replaced by the participant's mean for the relevant condition. Inspection of the distribution of responses indicated that some of the younger children and older adults failed to understand or comply with task instructions. Because failure to do so is shown by a flat distribution of long responses across all stimulus durations, participants were excluded if the difference in the proportion of long responses between any two stimuli was never greater than 0.6. Four five-year-olds and 1 eight-year-old were excluded from any further analyses according to this conservative criterion, as were 1 young-old and 2 old-old participants.

As with temporal generalization, data from the two samples were analyzed separately, although they are presented simultaneously. Figures 3a and 4a show the distribution of long responses across stimulus durations for each age group. Although all age groups produced orderly data, the S-shaped distributions were less steep for the youngest children. Two ANOVAs, each with a between-subjects variable of age group and a within-subject variable of stimulus duration, were used to examine the proportions of long responses given to each stimulus duration. There were no significant main effects of age, but the effects of stimulus duration were significant: $F(6, 654) = 736.71$, $MSE = 0.026$, $p < .001$, for the children and undergraduates; $F(6, 708) = 939.98$, $MSE = 0.025$, $p < .0001$, for the other adult groups. The interaction between age group and stimulus duration was significant only for the analysis of the children and undergraduates, $F(18, 654) = 5.08$, $p < .001$. Analysis of simple effects showed that age differences in the proportion of long responses were significant for the 200-, 300-, 400-, 600-, and 800-ms stimuli, $F(3, 109) > 4$, $p < .01$ for all comparisons, but not for the 500- and the 700-ms stimuli.

Further post hoc analyses showed that the only significant developmental differences in performance were between the 5-year-olds and the other age groups. The 5-year-olds gave significantly more long responses than all of the other age groups to the 200- and 300-ms stimuli and significantly fewer long responses to the 800-ms stimulus. They also gave more long responses to the 400-ms stimulus than did the 10-year-olds and undergraduates and significantly fewer long responses to the 600-ms stimulus than did the undergraduate group. In summary, there was a developmental change in the steepness of the gradient.

The bisection point for each participant was determined by linear interpolation of the two points between which it was known to lie (this is equivalent to reading the bisection point from the plot by eye for each participant, the method used by Wearden, Wearden, et al., 1997). As in Wearden, Wearden, et al.'s sample, there was a small number of participants for whom no clear bisection point could be determined (4 of the 5-year-olds). The mean bisection points for all other participants in the four age groups are shown in Table 2. There appeared to be a developmental trend for the bisection point to increase with age. However, ANOVAs found

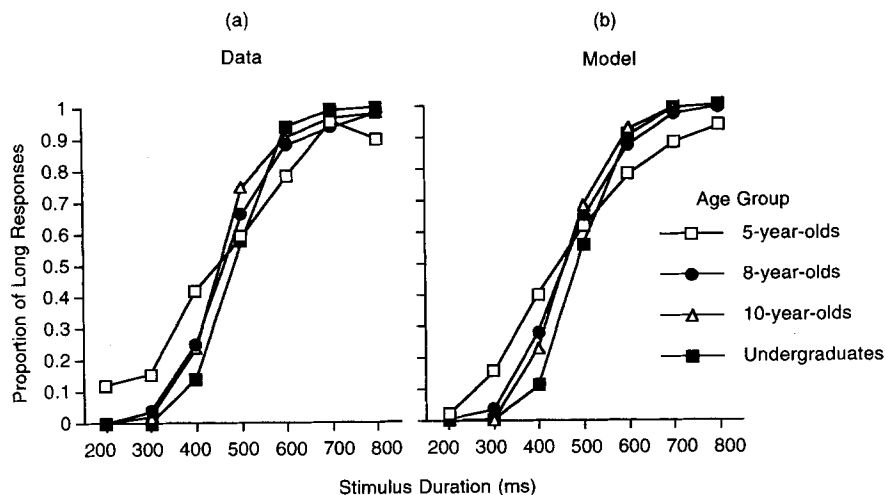


Figure 3. Proportion of long responses as a function of stimulus duration and age group for the children and undergraduates: (a) the data and (b) the model.

no significant main effect of age in either sample on bisection point: $F(3, 105) = 1.72$, $MSE = 4,474.20$, $p = .17$, for the children and undergraduates; $F(2, 118) < 1$, for the other adult groups.

For correlational analyses, temporal bisection performance of the young, young-old, and old-old adults was summarized by a single measure of the mean proportion correct (the proportion of short responses to stimulus durations of 200, 300, and 400 ms and the proportion of long responses to stimulus durations of 600, 700, and 800 ms; there was no correct response for 500 ms). In contrast to the results from the ANOVA, there was a significant correlation between age and performance, $r(119) = -.212$, $p < .05$, such that the mean proportion correct declined with increasing age. For the reduced sample with both temporal bisection data and background scores available ($n = 103$), this correlation between age and performance did not reach significance ($r = -.154$, $p > .1$), whereas the correlation between processing speed and perfor-

mance was significant ($r = .211$, $p < .05$). However, consistent with the results from the temporal generalization task, the weak correlation between age and performance was considerably reduced by partialing out the effect of processing speed ($r = .013$). Thus the small effect of age on temporal bisection is consistent with an age-related reduction in processing speed (Salthouse, 1996).

Discussion

The findings of this experiment were again consistent with those of Wearden, Wearden, et al. (1997). In their study, differences between 60- and 70-year-olds were significant on the temporal generalization task but not on the bisection task. Despite removing reminders of the long and short standards in our version of the task, which Wearden, Wearden, et al. suggested may increase the

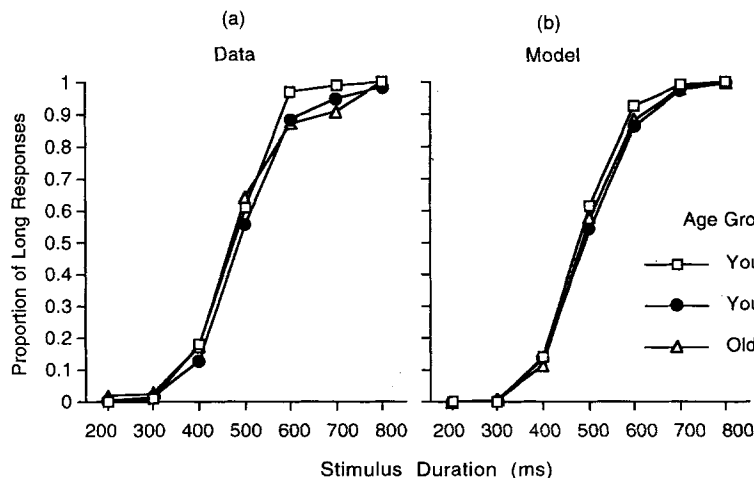


Figure 4. Proportion of long responses as a function of stimulus duration and age group for the young and older adults: (a) the data and (b) the model.

Table 2
Bisection Points for Each Age Group in Experiment 2

Group	Bisection point	
	<i>M</i>	<i>SD</i>
5-year-olds	444.2	97.8
8-year-olds	459.0	63.7
10-year-olds	458.6	54.2
Undergraduates	487.5	59.1
Young	476.5	55.6
Young-old	490.6	62.0
Old-old	482.4	75.4

task's sensitivity, we also found no differences between young and old adults. Age differences were only significant between 5-year-olds and older children and adults.

The general pattern of performance in the bisection task was also consistent with the findings of most previous studies using this task with human adults. As in previous studies, in all age groups the bisection points were close to the arithmetic mean. In fact, the bisection point in our undergraduate group was closer to the arithmetic mean than in Wearden, Wearden, et al.'s (1997) study using the same set of stimuli (487.5 compared with 440 ms). Removing the reminders of the standards may have made our bisection task more similar to Wearden and Ferrara's (1995) *partition task*, in which participants were not given any standards during the task but were simply told to judge whether tones were long or short. Under these circumstances, Wearden and Ferrara again found that the bisection point lies near the arithmetic mean, and in fact, inspection of their data shows that the point tends to lie even closer to the arithmetic mean in the partition task than in the standard bisection task. It is possible that in our version of the bisection task, as in the partition task, participants are not actually basing their responses on comparisons with memory representations of the standards, but rather extracting a central tendency for the stimulus set that enables them to assign approximately half of the stimuli as long and half as short (see Wearden & Ferrara, 1995, for a related suggestion). In our General Discussion, we develop a model of bisection performance based on this suggestion.

General Discussion

In Experiments 1 and 2, developmental differences in temporal generalization and bisection were examined at both ends of the life span. A large range of the life span was considered: The youngest participants were 5 years, and the oldest were between 75 and 99 years. Our findings show that it is possible to adapt these timing tasks successfully for use across such a wide range of the life span. Our results were on the whole consistent with existing findings. Developmental differences were more marked on the generalization task than on the bisection task, consistent with the findings of Wearden, Wearden, et al. (1997), as was our finding that all groups had bisection points close to the arithmetic mean.

One important general issue that the data from these studies enable us to address is whether developmental differences in timing behavior in childhood are qualitatively similar to developmental differences in aging. The results of the temporal generali-

zation task indicate that this is not the case. In our task, as in other studies, all adult groups produced generalization gradients with significant right asymmetry. By contrast, the children's generalization gradients showed left asymmetry, which was particularly marked in the youngest children. The difference in the shape of the generalization gradients between the children and the elderly adults (despite similar levels of overall performance: compare Figures 1a and 2a) is striking and needs to be explained by an account of timing behavior that can predict qualitatively different ways in which performance can change with development.

A Model of the Development of Biopsychological Time

We attempted to fit the main trends in the data using a model originally developed to account for serial-position effects in a range of perceptual identification and memory tasks (Neath, Brown, & Chater, 1998). The central idea of the model is that, in any perceptual identification task, test items are compared with memorial representations of the standard items. Thus the model is not specific to timing; it applies to any case in which items must be identified in terms of their position along a single dimension (such as weight, frequency, loudness, etc., as well as temporal duration). Standards, such as the 500-ms tone in our generalization task, are assumed to be represented veridically and in terms of the log transform of the stimulus duration (we make the pseudo-Fechnerian assumption that temporal durations are similar to other perceived magnitudes in that their representations on an internal psychological scale are proportional to the log-transformed value of the raw stimulus magnitude). In contrast to other recent models of timing, we assumed that there is no variability in the sampling of the memory representation of the standard. When a test tone is presented, however, we assumed that perception of its duration is susceptible to noise and that the amount of noise is proportional to the magnitude of the duration (in line with Weber's law). The central claim of the model is that the probability that a tone will be judged to be the same as the standard tone in temporal generalization will depend on the similarity between their representations. The similarity is defined as the inverse of the distance between them (i.e., $1/d$). Therefore, the closer the value of a perceived tone on the internal scale is to the value of the standard tone, the more likely is a yes response to result. More specifically, the probability of identifying a given tone as the standard will be proportional to the similarity of that tone to the standard.

This simple version of the model has just one free parameter when applied to adult data. This is the parameter that determines the amount of noise added to the perceived duration before it becomes represented on an internal psychological scale. More specifically, $R_i = T_i + q \cdot Y$, where Y is a normally distributed random number with $M = 0$ and $SD = 1$, q is the noise parameter, and R_i is the representation on the internal scale of a test tone T_i . Varying the noise parameter q , while holding other parameters constant, affects overall levels of performance, in a way similar to the variation of noise in other scalar timing models.

We required an additional parameter to account for what appears to be a developmental shift in memory representation of the standard. This is a distortion parameter, k , where k simply serves as a multiplier of the remembered standard. Thus if $k = 1$, the standard is remembered veridically, and if $k < 1$, then the standard is recalled as smaller than its true value. Varying the k parameter

(distortion) will affect the overall shape of the generalization gradient. The proportion of yes responses made to the standard tone gradually reduces with k , whereas the proportion of yes responses made to tones of shorter durations than the standard gradually increases. Eventually, when k has reduced sufficiently, the peak proportion of yes responses will be made to the tone shorter than the standard.

The model as described so far cannot make predictions on a trial-by-trial basis: To do so, it is necessary to incorporate a threshold-based decision rule into the model such that a yes response is made if the similarity between the test tone and the standard is greater than some threshold, which will be referred to as b , otherwise a no response is made (see Wearden, 1992). Increasing the value of the threshold leads to the generalization gradient becoming sharper and correspondingly more symmetrical. This is because higher thresholds effectively lead to more conservative performance: With a high threshold, a yes response will only be produced to perceived durations that are very similar to the memory representation of the standard duration.

Application of the Model to Temporal Generalization

The model was fitted to the temporal generalization data (Experiment 1) by choosing for each group values of the threshold, distortion, and noise parameters that minimized the overall summed squared error. The results are shown in Figures 1b and 2b, with Table 3 showing the parameter values that were chosen, along with the corresponding R^2 values. The two main developmental trends in the data (change in the steepness of the curves and a shift in the nature of the asymmetry) are both captured reasonably well by the model. It can be seen from Table 3 that, according to the model, the k parameter increases with development from a value of .87 (5-year-old children) to a value of 1.0 (all adult groups). Thus, the older participants do not show the systematic distortion of their memory representation of the standard that is evident in the younger children. The threshold parameter reduces systematically with age, being highest for the 5-year-olds and lowest for the old-old adults. By contrast, developmental changes with aging in the noise parameter are similar to those in childhood, with identical values of noise for the 5-year-old and old-old groups, and less noisy perceptual representations assumed to be available to young adults.

Application of the Model to Temporal Bisection

Consideration of the nature of the task facing participants in the temporal bisection experiments suggests two quite distinct ways in which the task might be undertaken. We refer to these as the *two-standards* method and the *partitioning* method (see Wearden & Ferrara, 1995).

The two-standards method. According to the two-standards assumption, participants maintain memory representations of the short and long standards. When a test tone is presented, it is compared with each of the standards in exactly the same way as test tones are compared with the single standard in the generalization task described earlier. Participants then respond long if the test tone is computed to be more similar to the long standard than to the short standard, and short otherwise. If such a model is implemented, changes in the noise parameter have the effect of flattening the response curve, in a manner similar to that observed in the data. However, in this version of the model, the mean bisection point falls at or very close to the geometric mean of the short and long standards. This follows directly from the logarithmic transformation of temporal durations assumed in the model. The question of the mechanisms underpinning the tendency for the bisection point to occur between the geometric and arithmetic means in human timing has received much attention in the literature (Wearden & Ferrara, 1995, 1996; Wearden, Rogers, et al., 1997). Our approach to this issue is to assume that bisection is done on the basis of one rather than two standards, as will now be described.

The partitioning method. According to the assumptions of what we term the partitioning method, the bisection task is performed by extracting a single central temporal duration. We assume that participants are sensitive to, and make use of, an implicit assumption that half of the responses should be short and half should be long. On the basis of this implicit assumption, participants infer an implicit mean that they use to partition their responses approximately equally in this way. Bisection judgments are then made by comparison with the inferred mean: A short response can be made if a given test tone is shorter than the mean, and a long response can be made if a given test tone is longer than the mean. Note that if the implicit mean used by participants were midway along the internal scale used to represent temporal durations, then the bisection point would occur exactly at the geometric mean (because of the logarithmic transformation that is assumed to

Table 3
Parameter Values Used in Modeling Generalization and Bisection Data

Group	Generalization				Bisection	
	Noise (q)	Threshold (b)	Distortion (k)	R^2	Noise (q)	R^2
5-year-olds	.13	30	.87	.988	.17	.975
8-year-olds	.12	25	.96	.936	.10	.995
10-year-olds	.08	15	.96	.997	.08	.994
Undergraduates	.07	11	1	.928	.07	.997
Young	.07	11	1	.991	.07	.998
Young-old	.085	9	1	.980	.08	.999
Old-old	.13	7	1	.921	.08	.990

occur). When the test durations are spaced equally in linear time, such an implicit mean would not serve to give equal numbers of long and short responses; to do this, the implicit mean must be the arithmetic mean.³

Because there is no straightforward way of determining the extent to which a given group of participants will be influenced by the hypothesized 50-50 response bias, we simply used the bisection point for a given group as a parameter in the model. Specifically, we assumed that the value of this parameter specified the mean that participants extracted, and that for each presented test duration participants estimated whether the duration was longer or shorter than this mean. This left just q free to vary, and values of q were chosen to give optimal fits that minimized overall summed squared error values (see Table 3). These fits are shown in Figures 3b and 4b, alongside the data. It can be seen that the model succeeds in capturing the main feature of the data: the gradual flattening of the response curve in the younger age groups and the small changes in the slope of the curve in the older adult groups.

Discussion and Comparison With Other Models

Interpretation of parameters. The noise parameter q changed in a similar way at both ends of the life span, showing a U-shaped pattern across development. For both tasks, its value decreased monotonically in modeling data from children of increasing age, and increased monotonically in modeling data from adult participants of increasing age. However, it can be seen from Table 3 that for four out of the seven groups, the values of the noise parameter that gave optimal fits were not identical across tasks, although they are of similar magnitude. We examined whether fits to the bisection data were significantly worse if the value of the noise parameter was indeed held constant across tasks. To do this, we tried modeling the bisection data using those values of the noise parameter that gave optimal fits to the generalization data. Maintaining noise values constant across tasks gave significantly worse fits in three out of the seven groups, $\chi^2(1) = 4.3$, $p < .05$, for the 5-year-olds; $\chi^2(1) = 4.37$, $p < .05$, for the young-old group; and $\chi^2(1) = 7.84$, $p < .01$, for the old-old group. Thus, although equally good fits to both the generalization and bisection data could be obtained for four out of the seven groups if values of the noise parameter were kept constant, for the youngest and oldest groups better fits could be obtained if the noise parameter varied. One possible explanation for such task differences in the value of the noise parameter is that, for these latter groups, the present data were not generated from identical sets of participants: We excluded more of the 5-year-olds' and older adults participants' data from the generalization than the bisection task. However, our findings are similar to those of Wearden, Wearden, et al. (1997), who also modeled the data of their older adult groups using different noise values for bisection than they used for generalization.

The amount of distortion (i.e., the extent to which k differed from 1) decreases in modeling the behavior of children of increasing age but remains constant at 1.0 in all adult groups. Because this aspect of the modeling underpins our claim that there is an important asymmetry between developmental improvement in children and developmental decline in the older adults, we examined whether it was really necessary to include the distortion parameter in fitting the children's data. To do this, we found the best possible

fits to the children's data that could be obtained by varying noise and threshold alone. These fits were then compared with the fits previously obtained by varying all three parameters. For data from all three groups of children, significantly better fits were obtained when distortion was varied with age than when it was fixed at a value of 1, $\chi^2(1) = 18.77$, $p < .001$, for the 5-year-olds; $\chi^2(1) = 4.64$, $p < .05$, for the 8-year-olds; and $\chi^2(1) = 16.75$, $p < .001$, for the 10-year-olds. Thus, it is necessary to vary distortion with age to account for the children's performance but not that of older adults (although it should be noted that our data are not sufficient to rule out other explanations for the children's left asymmetry, for example, in terms of some kind of response bias).

One way of interpreting the distortion parameter is as a measure of long-term forgetting of time intervals. As a result of forgetting, time intervals may be systematically misremembered as being shorter or longer than they actually were, with the former being the case in the present task. If children do consistently misremember time intervals in this way, then we might expect to see a greater tendency for "subjective shortening" (Wearden, Parry, & Stamp, 1998) in childhood, and systematic shortening in reproduction tasks in which a standard must be reproduced after a delay. We are not aware of any study that has measured subjective shortening in children; however, Block et al.'s (1999) meta-analysis of children's timing indicates that shortening in reproduction does indeed occur, at least under some circumstances.

A further asymmetry between children and the older adults is evident in the values of the threshold parameter, b . This is highest in the youngest group of participants and decreases monotonically to reach a minimum in the old-old participants. At first sight, this might be taken to suggest that the youngest participants are the most conservative (in that they require a higher level of similarity between test and standard before making a yes response) and that the oldest group are the least conservative. Such a conclusion may run counter to the normal expectations that older adults will be more conservative than younger adults (although empirical evidence for an increase in cautiousness in cognitive tasks with aging is equivocal; Salthouse, 1991) and that younger groups of children will be the least conservative. However, it is important to note that the threshold as implemented in the present model is not a simple response bias—any tendency to produce a greater or lesser proportion of yes responses overall can be overlaid onto the model as described here (which explicitly does not attempt to predict absolute level of yes responses produced, but only relative number). Thus the threshold parameter is not a straightforward measure of conservatism, and we do not view its decrease with age as evidence against previous claims of greater conservatism in responding in older participants.

Comparison with other models. Finally, we consider how the model we have proposed here differs from previous models of human timing. Perhaps the most fundamental issue concerns the transformations of temporal durations prior to their entry into psychological computations. The model that we have described contrasts with most models of human timing in that it assumes that

³ The idea that stimulus magnitudes are judged by comparing each stimulus with an internal mean has a long history (adaptation-level theory; Helson, 1964). Our account can be seen as a type of adaptation-level account.

perceived temporal durations undergo logarithmic transformation prior to the decision-making processes involved in temporal generalization and bisection. The logarithmic transformation is central to the model's production of asymmetrical temporal generalization gradients and to the model's temporal bisection at the geometric mean in the absence of any other bias. Thus, in contrast to most variants of scalar timing theory that invoke decision processes based on ratios of linear functions of time, the present model computes arithmetic differences of logarithmically transformed durations. Depending on the precise nature of other assumptions, such as the decision rules that are thought to operate, these procedures can of course turn out to be equivalent. Our preference for logarithmic transformation stems from our view that the current tasks are analogous to other perceptual identification tasks, and we have applied the model without alteration to verbal memory and pitch identification tasks. Logarithmic (or similar) transformation is widely assumed to occur in other perceptual dimensions, and in the absence of positive evidence suggesting that human timing is linear, it seems parsimonious to assume that duration is similar to other dimensions. We acknowledge, however, that the evidence for linear timing in animals may be more compelling.

A second feature of the model concerns the locus of noise in the model. We have assumed accurate memory of the standard durations and noisy perception of test durations. Previous models have typically assumed noise in the memory representations, sometimes in addition to noise in the test stimuli. In particular, Wearden, Wearden, et al. (1997) accounted for their findings of developmental change in temporal generalization in terms of an increase with age in the noise of memory representations of the standard. We have assumed veridical memory partly with the aim of reducing the number of free parameters as far as possible, and partly because the model we have presented is a development of a model of memory and perceptual identification in which there is little or no forgetting due to the passage of time per se (Neath et al., 1998). However, in practice, the behavior of the model is very similar if noise is added to the representation of the relevant standard tone in memory, and a smaller amount of noise is added to perceived test durations. At present, this choice seems to be one of aesthetic preference, at least with regard to the present data.

Last, our account of temporal bisection in terms of the partitioning of responses around a single standard differs from most previous accounts of bisection that assume comparison to two standards. The claim embodied in our model is that participants are sensitive to an implicit expectation that they should produce 50% long and 50% short responses. Wearden and Ferrara (1995) also suggested that bisection is done on the basis of a single standard, which they take to be the arithmetic mean. However, the present model differs from their bisection model in that we do not assume that the inferred mean is necessarily the arithmetic mean: Rather, responses will be partitioned on the basis of a standard that allows approximately half of all responses to be short and half to be long. Deviations from bisection at the geometric mean toward arithmetic mean bisection reflect sensitivity to this constraint. Furthermore, the tendency for children to have bisection points nearer to the geometric mean may reflect developmental changes in the sensitivity to this constraint.

Although our account contrasts with typical accounts of bisection, it makes sense of two puzzling findings in the literature. First, animals, in contrast to humans, would not be expected to show

sensitivity to such considerations, thus their bisection point falls at or very near the geometric mean (Church & Deluty, 1977). Second, it predicts that human participants should show a bisection point closer to the geometric mean under conditions in which geometric mean bisection would lead to 50% short and 50% long responding. Such conditions obtain when the durations are equally spaced on a logarithmic scale, and under such conditions participants do indeed show bisection closer to the geometric mean (Wearden & Ferrara, 1995; Wearden, Rogers, et al., 1997).⁴ Further studies varying the distribution of stimuli between the standards could be used to test whether the proposed 50-50 response bias can predict the bisection point better than other accounts. However, the main weakness with the present account is that the bisection point that supports such responding is taken as a parameter in the model, and providing an account of how such a point is extracted is beyond the scope of the current model (but see Helson, 1964).

Conclusion

We have shown that it is possible to use the temporal generalization and bisection procedures to examine timing across a wide range of the life span. These tasks yielded data that enable development to be considered within the theoretical approach that has been widely applied to animal and adult human timing. We have shown how a more general model of perceptual identification can be applied to such data. We note that no other model has been applied to the development of timing behavior across the life span and that other models of timing have, unlike the present model, in general been developed specifically to account for timing data.

These tasks measure developmental change only in certain aspects of timing behavior: namely, timing of very short unfilled intervals. Over longer time intervals, it is likely that numerous other cognitive processes, such as attention and inhibition, contribute to developmental changes (Block et al., 1998, 1999). However, we have shown that even on tasks that tap psychologically primitive timing operations, developmental improvement in performance in childhood cannot be seen as a mirror image of developmental change at the other end of the life span.

⁴ Allan and Gibbon (1991) found bisection points close to the geometric mean with both linear and logarithmic spacing. However, we note that their task instructions differed from those in other studies of human bisection in that participants were asked to judge whether a stimulus was the long or the short standard rather than judge which of the two standards it most resembled. Under these instructional conditions, the two-standards method may provide a more accurate characterization of performance.

References

- Allan, L. G., & Gibbon, J. (1991). Human bisection at the geometric mean. *Learning and Motivation*, 22, 39-58.
- Block, R. A., Zakay, D., & Hancock, P. A. (1998). Human aging and duration judgments: A meta-analytic review. *Psychology and Aging*, 13, 584-596.
- Block, R. A., Zakay, D., & Hancock, P. A. (1999). Developmental changes in human duration judgments: A meta-analytic review. *Developmental Review*, 19, 183-211.
- Brown, G. D. A., & Vousden, J. I. (1998). Adaptive analysis of sequential

- behaviour: Oscillators as rational mechanisms. In N. Chater & M. Oaksford (Eds.), *Rational models of cognition* (pp. 165–189). Oxford, England: Oxford University Press.
- Cattell, R. B., & Cattell, A. K. S. (1960). *Handbook for the Individual or Group Culture Fair Intelligence Test*. Champaign, IL: Institute for Personality and Ability Testing.
- Church, R. M. (1984). Properties of the internal clock. In J. Gibbon & L. G. Allan (Eds.), *Timing and time perception* (pp. 566–582). New York: New York Academy of Sciences.
- Church, R. M., & Deluty, M. Z. (1977). Bisection of temporal intervals. *Journal of Experimental Psychology: Animal Behavior Processes*, 3, 216–228.
- Church, R. M., & Gibbon, J. (1982). Temporal generalization. *Journal of Experimental Psychology: Animal Behavior Processes*, 8, 165–186.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84, 278–325.
- Gibbon, J. (1981). Two kinds of ambiguity in the study of psychological time. In M. L. Commons & J. A. Nevin (Eds.), *Quantitative analyses of behavior: Vol. 1. Discriminative properties of reinforcement schedules* (pp. 157–189). Cambridge, MA: Ballinger.
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. In J. Gibbon & L. G. Allan (Eds.), *Timing and time perception* (pp. 52–77). New York: New York Academy of Sciences.
- Heim, A. W. (1968). *AH4 Test*. Windsor, England: NFER-Nelson.
- Helson, H. (1964). *Adaptation-level theory*. New York: Harper & Row.
- Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107–129.
- Killeen, P. R., & Fetterman, J. G. (1988). A behavioral theory of timing. *Psychological Review*, 95, 274–295.
- Lejeune, H., Ferrara, A., Soffie, M., Bronchart, M., & Wearden, J. H. (1998). Peak performance in young adult and aged rats: Acquisition and adaptation to a changing temporal criterion. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 51B, 193–127.
- Maricq, A. V., Roberts, S., & Church, R. M. (1981). Methamphetamine and time estimation. *Journal of Experimental Psychology: Animal Behavior Processes*, 7, 18–30.
- Meck, W. H. (1983). Selective adjustment of the speed of the internal clock and memory processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 171–201.
- Meck, W. H. (1996). Neuropharmacology of timing and time perception. *Cognitive Brain Research*, 3, 227–242.
- Meck, W. H., Church, R. M., & Olton, D. S. (1984). Hippocampus, time and memory. *Behavioral Neuroscience*, 102, 54–60.
- Neath, I., Brown, G. D. A., & Chater, N. (1998). *Distinctiveness effects in perceptual identification*. Manuscript submitted for publication.
- Olton, D. S. (1989). Frontal cortex, timing and memory. *Neuropsychologia*, 27, 121–130.
- Penton-Voak, I. S., Edwards, H., Percival, A., & Wearden, J. H. (1996). Speeding up an internal clock in humans? Effects of click trains on subjective duration. *Journal of Experimental Psychology: Animal Behavior Processes*, 22, 307–320.
- Raslear, T. G. (1983). A test of the Pfanzagl bisection model in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 49–62.
- Raven, J. C., Raven, J., & Court, J. H. (1988). *The Mill Hill Vocabulary Scale*. London: H. K. Lewis.
- Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging*. Hillsdale, NJ: Erlbaum.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103, 403–428.
- Wearden, J. H. (1991a). Do humans possess an internal clock with scalar timing properties? *Learning and Motivation*, 22, 59–83.
- Wearden, J. H. (1991b). Human performance on an analogue of the interval bisection task. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 43B, 59–81.
- Wearden, J. H. (1992). Temporal generalization in humans. *Journal of Experimental Psychology: Animal Behavior Processes*, 18, 134–144.
- Wearden, J. H. (1994). Prescriptions for models of biopsychological time. In M. Oaksford & G. D. A. Brown (Eds.), *Neurodynamics and psychology* (pp. 215–236). London: Academic Press.
- Wearden, J. H. (1995). Categorical scaling of stimulus duration by humans. *Journal of Experimental Psychology: Animal Behavior Processes*, 21, 318–330.
- Wearden, J. H. (in press). Beyond the fields we know: Exploring and developing scalar timing theory. *Behavioural Processes*.
- Wearden, J. H., Denovan, L., Fakhri, M., & Haworth, R. (1997). Scalar timing in temporal generalization in humans with longer stimulus durations. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 502–511.
- Wearden, J. H., & Ferrara, A. (1995). Stimulus spacing effects in temporal bisection by humans. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 48B, 289–310.
- Wearden, J. H., & Ferrara, A. (1996). Stimulus range effects in temporal bisection by humans. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 49B, 24–44.
- Wearden, J. H., & Lejeune, H. (1993). Across the Great Divide: Animal psychology and time in humans. *Time and Society*, 2, 87–106.
- Wearden, J. H., & McShane, B. (1988). Interval production as an analogue of the peak procedure: Evidence for similarity of human and animal timing processes. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 40B, 363–375.
- Wearden, J. H., Parry, A., & Stamp, L. (1998). *Is subjective shortening in human memory unique to time representations?* Manuscript submitted for publication.
- Wearden, J. H., Rogers, P., & Thomas, R. (1997). Temporal bisection in humans with longer stimulus durations. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, 50B, 79–94.
- Wearden, J. H., Wearden, A. J., & Rabbitt, P. M. A. (1997). Age and IQ effects on stimulus and response timing. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 962–979.
- Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale—Revised*. New York: Psychological Corporation.

Received May 26, 1998

Revision received February 5, 1999

Accepted February 9, 1999 ■